SUT Journal of Mathematics
Vol. 59, No. 2 (2023), 91–116
DOI: 10.55937/sut/1699513820

# Parameter estimation and forecasts for an integrated Lee-Carter model

**Reo Kanazawa and Takeshi Kurosawa**

(Received August 6, 2023)

**Abstract.** The model to represent mortality that Lee and Carter proposed is still widely used. They proposed a method to estimate the parameters in the model with mortality data using the singular value decomposition. In addition, in the forecasting of future mortality, the time-dependent parameter is linearly estimated using ARIMA (0,1,0). This method treats the parameters as non-stochastic for part of the model construction using observed data, while it is stochastic for the forecasting of future mortality. This results in inconsistencies throughout the whole model. Girosi and King interpreted the parameters in the Lee-Carter model as random variables and provided an integrated expression for the two parts, which derived a single stochastic model; however, the covariance matrix in the single stochastic model is not the one by an ordinal ARIMA(0,1,0) and an estimation method was not clearly discussed, and estimating the parameters in their model appears to be difficult owing to the complicated covariance matrix. Therefore, this study proposes a new integrated model of the Lee-Carter model based on Girosi and King's interpretation. Our model is defined as a stochastic model that does not conflict with the concepts of Lee and Carter's existing model with the covariance matrix deduced by Girosi and King. Furthermore, we provide an estimation method for the parameters in our model by applying the idea of conditional distributions used in classical AR and MA models.

*AMS* 2020 *Mathematics Subject Classification.* 62P05

*Key words and phrases.* Lee-Carter model, mortality modeling, time series analysis, forecasts.

## §1. Introduction

We are exposed to various risks; thus, we need to take precautions against unexpected deaths, such as through natural disasters or illness. Insurance remedies the problem of these risks. With trillions of dollars per year in primary written premiums in the global insurance market, even small errors in an estimate of the premiums could have a significant impact on cash flow. This demonstrates the importance of designing and managing insurance products for each company, based on highly accurate predictions of future mortality. Lee and Carter [1] were the first to propose a practical method for stochastic modeling and mortality prediction. This model is a standard international method for estimating future mortalities. Although the model was proposed in 1992, it is still under development and has been applied in many countries around the world (e.g., see [2], [3], [4], and [5]). Tuljapurkar et al. [6] applied G7 data from 1950 to 1994 and found a linear decreasing in mortality rates as a universal pattern across countries.

In the parameter estimation by the Lee-Carter model, hereafter referred to as the LC model, various methods have been proposed, including the method using singular value decomposition (SVD) by Lee and Carter, such as the weighted least squares method, the maximum likelihood estimation method [7], and generalized linear models [8]. Some methods have been proposed for forecasting, such as fitting time-dependent parameters among the estimated parameters in ARIMA(0,1,0) and using a credibility approach to make predictions [9]. However, all methods focused independently on either the model construction part or the future forecasting part, resulting in inconsistent methods. Girosi and King [10] provided an integrated stochastic model based on the LC model, but did not provide a method for estimating its parameters because they were treated with complicated time-series models. In addition, their proposal is slightly inconsistent with the concept of the existing model; therefore, their interpretation of the parameters differs from that of the LC model. However, the slight difference is not intrinsic as they mentioned. The main problem is that they did not show a method to estimate parameters.

Therefore, this study proposes a new integrated model based on the idea of Girosi and King and its parameter estimation. Section 2 begins with an overview of the LC model and parameter estimation using SVD and forecasting methods with ARIMA(0,1,0) used by Lee and Carter. The integrated model by Girosi and King is introduced in Section 2.3. This integrates the model construction and future projection parts, and represents them as a single time-series model. Then, we propose a new integrated model following Girosi and King and discuss its estimation method in Section 3. Parameter estimation is based on the conditional likelihood function commonly used in AR and MR models, and we find that it is very easy to estimate because the solution can be obtained algebraically. Furthermore, the method proposed by Lee and Carter cannot estimate the variance of the error term in the model itself because the model construction part by Lee and Carter is constructed with a non-stochastic formulation, but our method provides an estimator of the variance. In Section 4, the bias and MSE of the proposed estimates and the accuracy of future mortality are validated using Monte Carlo simulations. The results show that our proposal is accurate and can be computed faster than the optimization function in R. Finally, in Sections 5 and 6, we apply our model to real data and show that the estimation result does not conflict with the LC model.

## §2. Lee-Carter Model

The LC model [1], hereafter referred to as the LC model, is a model of mortality $Q_{x,t}$ for age $x$ at time $t$ in the form of

$$M_{x,t} := \log Q_{x,t} = a_x + b_x k_t + \varepsilon_{x,t}, \tag{2.1}$$

where $\varepsilon_{x,t}$ is an error term with a mean of 0. The definitions of each parameter are as follows:

$a_x \cdots$ Logarithm of average mortality at each age $x$,

$k_t \cdots$ Deviation of the average mortality $a_x$ at each time point $t$,

$b_x \cdots$ The degree to which mortality is affected at each age corresponding to $k_t$.

Let the vectors $\boldsymbol{a}, \boldsymbol{b}$, $\boldsymbol{M}_t$, and $\boldsymbol{\varepsilon}_t$ be as follows:
(2.2)

$$\boldsymbol{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}, \quad \boldsymbol{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}, \quad \boldsymbol{M}_t = \begin{pmatrix} \log Q_{1,t} \\ \vdots \\ \log Q_{n,t} \end{pmatrix}, \quad \boldsymbol{\varepsilon}_t = \begin{pmatrix} \varepsilon_{1,t} \\ \vdots \\ \varepsilon_{n,t} \end{pmatrix},$$

where $n$ denotes the final age of the life table.

### 2.1. A parameter estimation method by Lee and Carter

Lee and Carter proposed a method for estimating parameters $a_x, b_x$ and $k_t$. This method is based on SVD. The parameters were determined by minimizing

$$\sum_{x=1}^{n}\sum_{t=1}^{T}(\log q_{x,t} - a_x - b_x k_t)^2,$$

which is the minimization of the sum of the error terms $\varepsilon_{x,t}$ under the parameter constraint,

$$\sum_{x=1}^{n} b_x = 1, \qquad \sum_{t=1}^{T} k_t = 0 \tag{2.3}$$

using actual mortality data $q_{x,t}$ of $Q_{x,t}$ for $T$ years. Under this constraint, they used SVD and estimated the parameters.

### 2.2. A prediction of future derivation of the average mortality

Lee and Carter applied their model to the U.S. mortality rate and observed that the estimates $\hat{k}_t$ $(1 \leq t \leq T)$ of the average mortality rates are approximately linear. They assumed that the trend in the derivation of the average mortality rate continued. Thus, they formulate $k_t$ as follows:

$$k_t = k_{t-1} + \theta + \xi_t \tag{2.4}$$

at each time $t$ in the future, where $\xi_t$ is an error term with mean 0. This was a random walk model with ARIMA(0,1,0). In other words, $k_t$ is treated as a stochastic model for future mortality prediction at time $t$ $(t > T)$. In the model construction stage in (2.1), $k_t$ is treated as a parameter in the model. $k_t$ is a random walk model with ARIMA(0,1,0) at the stage of future derivation of the average mortality rate in (2.4), which is inconsistent and completely distinct between the model construction stage and future prediction. It is appropriate to distinguish between parameters and random variables in the expression of $k_t$. It would be convenient to use another notation $\kappa_t$ instead of $k_t$ in (2.4) to stress the randomness of the model for the expression of the future derivation of average mortality.

### 2.3. A representation of the integrated model by Girosi and King

To integrate the LC model, we have

$$\begin{aligned} \boldsymbol{M}_t &= \boldsymbol{a} + \kappa_t \boldsymbol{b} + \boldsymbol{\varepsilon}_t, \\ \kappa_t &= \kappa_{t-1} + \theta + \xi_t \end{aligned} \tag{2.5}$$

with (2.2). While Lee and Carter used the stochastic model ARIMA(0,1,0) to represent the derivation of the average mortality in the forecasting model, as in (2.4), they treated it as a non-stochastic model in (2.1) at the model construction stage. The attempt of the integration of the two parts and the interpretation of the LC model was deeply discussed by Girosi and King [10]. To obtain an integrated expression for the model construction and forecasting parts of the model, Girosi and King formulated (2.5) using the stochastic variable $\kappa_t$ in the model construction part. In fact, they express the LC model instead of (2.5) as

$$\begin{aligned} \boldsymbol{M}_t &= \overline{\boldsymbol{m}} + \kappa_t \boldsymbol{b} + \boldsymbol{\varepsilon}_t, \\ \kappa_t &= \kappa_{t-1} + \theta + \xi_t \end{aligned} \tag{2.6}$$

by setting

$$\hat{\boldsymbol{a}} = \overline{\boldsymbol{m}} = \begin{pmatrix} \frac{1}{T}(\log q_{1,1} + \cdots + \log q_{1,T}) \\ \vdots \\ \frac{1}{T}(\log q_{n,1} + \cdots + \log q_{n,T}) \end{pmatrix}, \tag{2.7}$$

where the vector $\boldsymbol{\varepsilon}_t$ and scalar $\xi_t$ are assumed to have normal errors with variances $\sigma_\varepsilon^2 I_n$ and $\sigma_\xi^2 > 0$. In other words, the vector $\boldsymbol{\varepsilon}_t$ has mutually independent error terms $\varepsilon_{1,t}, \varepsilon_{2,t}, \ldots, \varepsilon_{n,t}$ and

$$\boldsymbol{\varepsilon}_t \sim \mathcal{N}_n(\mathbf{0}, \sigma_\varepsilon^2 I_n), \quad \xi_t \sim \mathcal{N}(0, \sigma_\xi^2).$$

Therefore, $\kappa_t$ in (2.5) is typically treated as a random variable. The two equations can be unified into a single equation by vanishing $\kappa_t$ in (2.5), as follows:

$$\boldsymbol{M}_t = \boldsymbol{M}_{t-1} + \theta \boldsymbol{b} + (\xi_t \boldsymbol{b} + \boldsymbol{\varepsilon}_t - \boldsymbol{\varepsilon}_{t-1}).$$

Furthermore, they set $\boldsymbol{b}$ and $\theta$ as follows:

$$\theta := \theta_{\mathrm{gk}} = \|\boldsymbol{\psi}\|, \quad \boldsymbol{b} := \boldsymbol{b}_{\mathrm{gk}} = \frac{\boldsymbol{\psi}}{\|\boldsymbol{\psi}\|}, \tag{2.8}$$

we obtain $\theta_{\mathrm{gk}} \boldsymbol{b}_{\mathrm{gk}} = \boldsymbol{\psi}$ and

$$\boldsymbol{M}_t = \boldsymbol{M}_{t-1} + \boldsymbol{\psi} + \left( \frac{\boldsymbol{\psi}}{\|\boldsymbol{\psi}\|} \xi_t + \boldsymbol{\varepsilon}_t - \boldsymbol{\varepsilon}_{t-1} \right). \tag{2.9}$$

Note that the constraint of (2.3) is not satisfied when $\boldsymbol{b}$ is defined by (2.8). This means that even if $\boldsymbol{\psi}$ is estimated correctly, it does not coincide with Lee and Carter's estimate using SVD. However, as they mentioned, this is not an intrinsic problem because the parameter can be adjusted after the estimation. Thus, they obtained the single random walk model ARIMA(0,1,0). However, the biggest difference between ordinal ARIMA(0,1,0) and (2.9) is error structure. The error distribution in ARIMA(0,1,0) is $\boldsymbol{\varepsilon}_t$, and then the covariance matrix by (2.9) is different and includes $\boldsymbol{\psi}$ in the covariance matrix. They did not show an estimation method for the above single random walk (2.9).

By setting $\boldsymbol{Y}_t = \boldsymbol{M}_{t+1} - \boldsymbol{M}_t$, we obtain

$$\boldsymbol{Y}_t \sim \mathcal{N}_n(\boldsymbol{\psi}, \Sigma_{\mathrm{gk}}),$$

where

$$\boldsymbol{Y}_t = \begin{pmatrix} Y_{1,t} \\ \vdots \\ Y_{n,t} \end{pmatrix}, \quad \Sigma_{\mathrm{gk}} = \sigma_\xi^2 \frac{\boldsymbol{\psi}\boldsymbol{\psi}'}{\|\boldsymbol{\psi}\|^2} + 2\sigma_\varepsilon^2 I_n.$$

Generally, if $A$ is a matrix of type $n \times m$ and $B$ is a matrix of type $m \times n$, then

$$(I_n + AB)^{-1} = I_n - A(I_m + BA)^{-1}B.$$

Therefore,

$$\Sigma_{\mathrm{gk}}^{-1} = \frac{1}{2\sigma_\varepsilon^2}\left(I_n - \frac{\sigma_\xi^2}{(2\sigma_\varepsilon^2 + \sigma_\xi^2)\|\boldsymbol{\psi}\|^2}\boldsymbol{\psi}\boldsymbol{\psi}'\right).$$

For a time difference of 1, we have

$$\mathrm{Cov}(\boldsymbol{Y}_t, \boldsymbol{Y}_{t-1}) = -\sigma_\varepsilon^2 I_n$$

and there is no correlation for a time difference of 2 or more. That is,

$$\mathrm{Cov}(\boldsymbol{Y}_t, \boldsymbol{Y}_{t-s}) = \boldsymbol{O} \qquad (|s| \geq 2).$$

## §3. A new integrated model and parameter estimation

### 3.1. A new integrated model

Girosi and King set $\theta$ and $\boldsymbol{b}$ as in (2.8). However, the definition loses affinity for the original LC model because the estimated result with SVD by Lee and Carter is different from the estimate of $\boldsymbol{b}$ defined by Girosi and King because the LC model has the constraints in (2.3). Therefore, we define

$$\theta = \sum_{x=1}^{n} \psi_x, \quad \boldsymbol{b} = \frac{\boldsymbol{\psi}}{\sum_{x=1}^{n} \psi_x} \tag{3.1}$$

to hold the affinity for the estimate obtained by SVD because

$$\sum_{x=1}^{n} b_x = \sum_{x=1}^{n} \frac{\psi_x}{\sum_{i=1}^{n} \psi_i} = 1, \tag{3.2}$$

which is the constraint on $\boldsymbol{b}$ in (2.3). Next, we check the constraints on $k_t$ $(1 \leq t \leq T)$. Here, $k_t$ is interpreted as the (ensemble) expected value of the random variable $\kappa_t$ with

$$E[\kappa_t] = k_t.$$

We also define $\overline{\boldsymbol{M}}$ as

$$\overline{\boldsymbol{M}} = \begin{pmatrix} \frac{1}{T}(\log Q_{1,1} + \cdots + \log Q_{1,T}) \\ \vdots \\ \frac{1}{T}(\log Q_{n,1} + \cdots + \log Q_{n,T}) \end{pmatrix}$$

and let $\overline{M}_x$ be its components. Because we assume $\overline{\boldsymbol{M}}$ is a vector of random variables, we set its observed value to $\overline{\boldsymbol{m}}$ in (2.7). Thus, we have the following formula:

$$\boldsymbol{M}_t = \overline{\boldsymbol{M}} + \kappa_t \boldsymbol{b} + \boldsymbol{\varepsilon}_t.$$

In this case, using constraint (3.2) on $b_x$, we obtain, by (2.1) with the estimate of $\boldsymbol{a}$ in (2.7),

$$\sum_{x=1}^{n} E\left[\log Q_{x,t}\right] = \sum_{x=1}^{n} E\left[\overline{M}_x + \kappa_t b_x + \varepsilon_{x,t}\right] = \sum_{x=1}^{n} E[\overline{M}_x] + k_t.$$

Therefore, we obtain

$$k_t = \sum_{x=1}^{n} \left( E[\log Q_{x,t}] - E[\overline{M}_x] \right). \tag{3.3}$$

Then, the summation of both sides of (3.3) running $t$ from 1 to $T$ yields

$$\sum_{t=1}^{T} k_t = \sum_{t=1}^{T} \sum_{x=1}^{n} E[\log Q_{x,t}] - T \sum_{x=1}^{n} E[\overline{M}_x] = T \sum_{x=1}^{n} E[\overline{M}_x] - T \sum_{x=1}^{n} E[\overline{M}_x] = 0.$$

Therefore, $k_t$ is defined by the expected value of $\kappa_t$ also satisfies the constraint condition in (2.3).

Moreover, Girosi and King assume $\xi_t$ independently with respect to time $t$ in (2.5). However, we may assume a general assumption on $\xi_t$ that they are not always independent, but we restrict the form of $\xi_t$ to deduce a quasi-likelihood function defined below to be expressed as a product of conditional distributions, which is a well-known idea to obtain a quasi-likelihood function in the classical AR and MA models. Thus, we consider error terms $\xi_1, \xi_2, \ldots, \xi_t$ of the form

$$\xi_t = \zeta_t - \zeta_{t-1}, \quad \zeta_t \sim \mathcal{N}(0, \sigma_\zeta^2) \tag{3.4}$$

with $\zeta_0, \zeta_1, \ldots, \zeta_t$ independent for each $t$, which are independent from $\boldsymbol{\varepsilon}_t$. Thus, $\xi_t$ is no longer an independent sequence with respect to time $t$. Thus, the two-stage model of (2.5) is expressed as

$$\begin{aligned} \boldsymbol{M}_t &= \boldsymbol{a} + \kappa_t \boldsymbol{b} + \boldsymbol{\varepsilon}_t, \\ \kappa_t &= \kappa_{t-1} + \theta + \zeta_t - \zeta_{t-1} \end{aligned} \tag{3.5}$$

and $k_t = E(\kappa_t)$ with constraints (2.3). This is our proposed model. By expressing this with $l_t = \kappa_t - \zeta_t$, we have

$$l_t = l_{t-1} + \theta.$$

This time-dependent relationship can be interpreted as random walk. Then, using (3.1), we have that

$$\boldsymbol{M}_{t+1} = \boldsymbol{M}_t + \boldsymbol{\psi} + \frac{\boldsymbol{\psi}}{\sum_{x=1}^{n} \psi_x}(\zeta_{t+1} - \zeta_t) + \boldsymbol{\varepsilon}_{t+1} - \boldsymbol{\varepsilon}_t. \tag{3.6}$$

Let $\boldsymbol{Y}_t = \boldsymbol{M}_{t+1} - \boldsymbol{M}_t$. Then, we have

$$\boldsymbol{Y}_t \sim \mathcal{N}_n\left(\boldsymbol{\psi}, 2\Sigma\right), \tag{3.7}$$

where

$$\Sigma = \sigma_\zeta^2 \frac{\boldsymbol{\psi}\boldsymbol{\psi}'}{\left(\sum_{x=1}^{n} \psi_x\right)^2} + \sigma_\varepsilon^2 I_n. \tag{3.8}$$

Note that

$$\begin{aligned} &\mathrm{Cov}(\boldsymbol{Y}_t, \boldsymbol{Y}_{t-1}) = -\Sigma, \\ &\mathrm{Cov}(\boldsymbol{Y}_t, \boldsymbol{Y}_{t-s}) = \boldsymbol{O} \qquad (|s| \geq 2) \end{aligned} \tag{3.9}$$

for a time difference of 1 and for a time difference of 2 or more, respectively.

### 3.2. Quasi-Likelihood function

In this subsection, we discuss the parameter estimation method of the new integrated model. The model resembles that proposed by Girosi and King. Therefore, parameter estimation is not easy, as it is not. However, our formulation provides a quasi-likelihood function with additional assumptions in the model. Assume that

$$\zeta_1 = 0, \quad \boldsymbol{\varepsilon}_1 = \boldsymbol{0} \tag{3.10}$$

and consider quasi-maximum likelihood estimation using a conditional likelihood function. Using the assumption of the error term, the conditional distribution for

$$\boldsymbol{Y}_1 = \boldsymbol{\psi} + \frac{\boldsymbol{\psi}}{\sum_{x=1}^{n} \psi_x}(\zeta_2 - \zeta_1) + \boldsymbol{\varepsilon}_2 - \boldsymbol{\varepsilon}_1 = \boldsymbol{\psi} + \frac{\boldsymbol{\psi}}{\sum_{x=1}^{n} \psi_x}\zeta_2 + \boldsymbol{\varepsilon}_2 \tag{3.11}$$

leads to

$$\boldsymbol{Y}_1 | (\zeta_1 = 0, \boldsymbol{\varepsilon}_1 = \boldsymbol{0}) \sim \mathcal{N}_n\left(\boldsymbol{\psi}, \Sigma\right).$$

Thus, we obtain the following probability density function:

$$f_{\boldsymbol{Y}_1|\zeta_1,\boldsymbol{\varepsilon}_1}(\boldsymbol{y}_1 | \zeta_1 = 0, \boldsymbol{\varepsilon}_1 = \boldsymbol{0}) = \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{\det \Sigma}} \exp\left\{ -\frac{1}{2}(\boldsymbol{y}_1 - \boldsymbol{\psi})'\Sigma^{-1}(\boldsymbol{y}_1 - \boldsymbol{\psi}) \right\}.$$

In addition, using the same arguments as those in the previous section, we have

$$\Sigma^{-1} = \frac{1}{\sigma_\varepsilon^2}\left(I_n - \frac{\sigma_\zeta^2}{\sigma_\zeta^2\|\boldsymbol{\psi}\|^2 + \sigma_\varepsilon^2\left(\sum_{i=1}^n \psi_i\right)^2}\boldsymbol{\psi}\boldsymbol{\psi}'\right). \tag{3.12}$$

Under the condition that $\boldsymbol{Y}_1 = \boldsymbol{y}_1$ is observed, we consider the distribution of $\boldsymbol{Y}_2$. The conditional distribution $\boldsymbol{Y}_2$ given $\boldsymbol{Y}_1 = \boldsymbol{y}_1$ is

$$\boldsymbol{Y}_2 = \boldsymbol{\psi} + \left(\frac{\boldsymbol{\psi}}{\sum_{x=1}^n \psi_x}(\zeta_3 - \zeta_2) + \boldsymbol{\varepsilon}_3 - \boldsymbol{\varepsilon}_2\right),$$

we obtain the following by (3.11):

$$\boldsymbol{Y}_2 = 2\boldsymbol{\psi} - \boldsymbol{y}_1 + \left(\frac{\boldsymbol{\psi}}{\sum_{x=1}^n \psi_x}\zeta_3 + \boldsymbol{\varepsilon}_3\right).$$

Thus, we get

$$\boldsymbol{Y}_2 | (\boldsymbol{Y}_1 = \boldsymbol{y}_1, \zeta_1 = 0, \boldsymbol{\varepsilon}_1 = \boldsymbol{0}) \sim \mathcal{N}_n\left(2\boldsymbol{\psi} - \boldsymbol{y}_1, \Sigma\right).$$

Thus, we have the following conditional density function:

$$\begin{aligned}
&f_{\boldsymbol{Y}_2|\boldsymbol{Y}_1,\zeta_1,\boldsymbol{\varepsilon}_1}(\boldsymbol{y}_2|\boldsymbol{y}_1, \zeta_1 = 0, \boldsymbol{\varepsilon}_1 = \boldsymbol{0}) \\
&\quad = \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{\det\Sigma}}\exp\left\{-\frac{1}{2}(\boldsymbol{y}_2 - 2\boldsymbol{\psi} + \boldsymbol{y}_1)'\Sigma^{-1}(\boldsymbol{y}_2 - 2\boldsymbol{\psi} + \boldsymbol{y}_1)\right\}.
\end{aligned}$$

Similarly, the distribution of $\boldsymbol{Y}_t$ under $\boldsymbol{Y}_1 = \boldsymbol{y}_1, \ldots, \boldsymbol{Y}_{t-1} = \boldsymbol{y}_{t-1}$ can be rearranged by summing up from

$$\begin{aligned}
\boldsymbol{y}_k &= \boldsymbol{\psi} + \left(\frac{\boldsymbol{\psi}}{\sum_{x=1}^n \psi_x}(\zeta_{k+1} - \zeta_k) + \boldsymbol{\varepsilon}_{k+1} - \boldsymbol{\varepsilon}_k\right) \qquad (k = 1, 2, \ldots, t-1), \\
\boldsymbol{Y}_t &= \boldsymbol{\psi} + \left(\frac{\boldsymbol{\psi}}{\sum_{x=1}^n \psi_x}(\zeta_{t+1} - \zeta_t) + \boldsymbol{\varepsilon}_{t+1} - \boldsymbol{\varepsilon}_t\right)
\end{aligned}$$

as

$$\boldsymbol{Y}_t = t\boldsymbol{\psi} - \sum_{k=1}^{t-1}\boldsymbol{y}_k + \left(\frac{\boldsymbol{\psi}}{\sum_{x=1}^n \psi_x}(\zeta_{t+1} - \zeta_1) + \boldsymbol{\varepsilon}_{t+1} - \boldsymbol{\varepsilon}_1\right).$$

Therefore, from (3.10), we obtain

$$\boldsymbol{Y}_t | (\boldsymbol{Y}_{t-1} = \boldsymbol{y}_{t-1}, \ldots, \boldsymbol{Y}_1 = \boldsymbol{y}_1, \zeta_1 = 0, \boldsymbol{\varepsilon}_1 = \boldsymbol{0}) \sim \mathcal{N}_n\left(t\boldsymbol{\psi} - \sum_{k=1}^{t-1}\boldsymbol{y}_k, \Sigma\right).$$

Thus, the conditional density function is

$$
\begin{aligned}
&f_{\boldsymbol{Y}_t|\boldsymbol{Y}_{t-1},\ldots,\boldsymbol{Y}_1\zeta_1,\boldsymbol{\varepsilon}_1}(\boldsymbol{y}_t|\boldsymbol{y}_{t-1},\ldots,\boldsymbol{y}_1,\zeta_1=0,\boldsymbol{\varepsilon}_1=\mathbf{0})\\
&=\frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{\det\Sigma}}\exp\left\{-\frac{1}{2}(\boldsymbol{y}_t-t\boldsymbol{\psi}+\sum_{k=1}^{t-1}\boldsymbol{y}_k)'\Sigma^{-1}(\boldsymbol{y}_t-t\boldsymbol{\psi}+\sum_{k=1}^{t-1}\boldsymbol{y}_k)\right\}.
\end{aligned}
$$

By multiplying the individual density functions, the joint likelihood is given by

$$
\begin{aligned}
&f_{\boldsymbol{Y}_{T-1},\boldsymbol{Y}_{T-2},\ldots,\boldsymbol{Y}_1|\zeta_1,\boldsymbol{\varepsilon}_1}\left(\boldsymbol{y}_{T-1},\boldsymbol{y}_{T-2},\ldots,\boldsymbol{y}_1\mid\zeta_1=0,\boldsymbol{\varepsilon}_1=\mathbf{0}\right)\\
&=f_{\boldsymbol{Y}_1|\zeta_1,\boldsymbol{\varepsilon}_1}\left(\boldsymbol{y}_1\mid\zeta_1=0,\boldsymbol{\varepsilon}_1=\mathbf{0}\right)\\
&\quad\times\prod_{t=2}^{T-1}f_{\boldsymbol{Y}_t|\boldsymbol{Y}_{t-1},\boldsymbol{Y}_{t-2},\ldots,\boldsymbol{Y}_1,\zeta_1,\boldsymbol{\varepsilon}_1}\left(\boldsymbol{y}_t\mid\boldsymbol{y}_{t-1},\boldsymbol{y}_{t-2},\ldots,\boldsymbol{y}_1,\zeta_1=0,\boldsymbol{\varepsilon}_1=\mathbf{0}\right),
\end{aligned}
$$

where $T$ denotes the number of vectors observed. The following quasi-likelihood function is then obtained:

$$
\begin{aligned}
\mathscr{L}(\boldsymbol{\psi},\sigma_\varepsilon^2,\sigma_\zeta^2)=&-\frac{(T-1)n}{2}\log(2\pi)+\frac{T-1}{2}\log(\det\Sigma^{-1})\\
&-\frac{1}{2}\sum_{i=1}^{T-1}\left(\sum_{k=1}^{i}\boldsymbol{y}_k-i\boldsymbol{\psi}\right)'\Sigma^{-1}\left(\sum_{k=1}^{i}\boldsymbol{y}_k-i\boldsymbol{\psi}\right).
\end{aligned}
\tag{3.13}
$$

Here, we note that $\Sigma$ is a function of $\sigma_\varepsilon^2,\sigma_\zeta^2$, and $\boldsymbol{\psi}$ as in (3.8). The biggest difficulty is that the mean parameter $\boldsymbol{\psi}$ is included in the covariance matrix $\Sigma$.

### 3.3. Parameter estimations and the bias

Next, we are interested in the parameter estimation of $\sigma_\varepsilon^2,\sigma_\zeta^2$, and $\boldsymbol{\psi}$. The use of optimal functions in the statistical package enabled us to estimate these parameters. However, we discuss an algebraical closed estimate of these parameters. To get the algebraical solution, we propose a two-step estimation method to obtain an algebraic solution. First, we do not treat $\Sigma$ as a function of $\boldsymbol{\psi}$, although $\Sigma$ is a function of $\boldsymbol{\psi}$. This implies that we estimate $\boldsymbol{\psi}$ and $\Sigma$ independently in the first stage and then substitute the estimated value of $\boldsymbol{\psi}$ into $\Sigma$. Next, we estimate $\sigma_\varepsilon^2$ and $\sigma_\zeta^2$ using the estimate $\hat{\Sigma}$ with the estimate $\hat{\boldsymbol{\psi}}$. The estimates of $\sigma_\varepsilon^2$ and $\sigma_\zeta^2$ are given by (3.19) and (3.20), respectively. Before describing the estimates of $\sigma_\varepsilon^2$ and $\sigma_\zeta^2$, we consider the estimates of $\Sigma$ and $\boldsymbol{\psi}$.

As the distribution follows (3.7), we may consider

$$\begin{aligned}\hat{\boldsymbol{\psi}}_1 &= \frac{1}{T-1}\sum_{i=1}^{T-1}\boldsymbol{y}_i = \frac{\boldsymbol{m}_T-\boldsymbol{m}_1}{T-1},\\ \hat{\Sigma}_1 &= \frac{1}{2(T-1)}\sum_{i=1}^{T-1}\left(\boldsymbol{y}_k-\hat{\boldsymbol{\psi}}_1\right)\left(\boldsymbol{y}_k-\hat{\boldsymbol{\psi}}_1\right)'\end{aligned} \tag{3.14}$$

as the intuitive estimator. The estimator $\hat{\boldsymbol{\psi}}_1$ is used in ARIMA(0,1,0) as noted in [10, Sect. 3.1].

Next, we consider the estimates of $\boldsymbol{\psi}$ and $\Sigma$ using the quasi-likelihood function as an estimator different from $\hat{\boldsymbol{\psi}}_1$ and $\hat{\Sigma}_1$. Here, $\mathscr{L}(\boldsymbol{\psi},\sigma_\varepsilon^2,\sigma_\zeta^2)$ is a function of $\sigma_\varepsilon^2,\sigma_\zeta^2$, and $\boldsymbol{\psi}$. However, we consider $\mathscr{L}(\boldsymbol{\psi},\Sigma)$ as a function of $\boldsymbol{\psi}$ and $\Sigma$ instead of $\mathscr{L}(\boldsymbol{\psi},\sigma_\varepsilon^2,\sigma_\zeta^2)$ and find the maximum solution of $\mathscr{L}(\boldsymbol{\psi},\Sigma)$. In other words, although $\Sigma$ is a function of $\boldsymbol{\psi}$, we do not treat it as a function of $\boldsymbol{\psi}$. Then, we perform a two-step estimation for $\sigma_\varepsilon^2$ and $\sigma_\zeta^2$ after obtaining the estimates of $\boldsymbol{\psi}$ and $\Sigma$. Solving

$$\begin{aligned}\frac{\partial}{\partial\boldsymbol{\psi}}\mathscr{L}(\boldsymbol{\psi},\Sigma) &= -\frac{\partial}{\partial\boldsymbol{\psi}}\frac{1}{2}\sum_{i=1}^{T-1}\left(\sum_{k=1}^{i}\boldsymbol{y}_k-i\boldsymbol{\psi}\right)'\Sigma^{-1}\left(\sum_{k=1}^{i}\boldsymbol{y}_k-i\boldsymbol{\psi}\right)\\ &= \sum_{i=1}^{T-1}i\Sigma^{-1}\left(\sum_{k=1}^{i}\boldsymbol{y}_k\right)-\sum_{i=1}^{T-1}i^2\Sigma^{-1}\boldsymbol{\psi}=\boldsymbol{0},\end{aligned}$$

we get

$$\hat{\boldsymbol{\psi}}_2 = \frac{\sum_{i=1}^{T-1}\sum_{k=1}^{i}i\boldsymbol{y}_k}{\sum_{i=1}^{T-1}i^2} = \frac{\sum_{k=1}^{T-1}\sum_{i=k}^{T-1}i\boldsymbol{y}_k}{\sum_{i=1}^{T-1}i^2} = \frac{3\sum_{k=1}^{T-1}(k+T-1)(T-k)\boldsymbol{y}_k}{T(T-1)(2T-1)}. \tag{3.15}$$

Next, we calculate the quasi-maximum likelihood estimator of $\Sigma$. Differentiating the likelihood function above by $\Sigma^{-1}$ and solving

$$\begin{aligned}&\frac{\partial}{\partial\Sigma^{-1}}\mathscr{L}(\boldsymbol{\psi},\Sigma)\\ &= \frac{\partial}{\partial\Sigma^{-1}}\left[\frac{T-1}{2}\log(\det\Sigma^{-1})-\frac{1}{2}\sum_{i=1}^{T-1}\left(\sum_{k=1}^{i}\boldsymbol{y}_k-i\boldsymbol{\psi}\right)'\Sigma^{-1}\left(\sum_{k=1}^{i}\boldsymbol{y}_k-i\boldsymbol{\psi}\right)\right]\\ &= \frac{T-1}{2}\Sigma-\frac{1}{2}\sum_{i=1}^{T-1}\left(\sum_{k=1}^{i}\boldsymbol{y}_k-i\boldsymbol{\psi}\right)\left(\sum_{k=1}^{i}\boldsymbol{y}_k-i\boldsymbol{\psi}\right)'=\boldsymbol{O},\end{aligned}$$

we obtain

$$\hat{\Sigma}_2 = \frac{1}{T-1}\sum_{i=1}^{T-1}\left(\sum_{k=1}^{i}\boldsymbol{y}_k - i\hat{\boldsymbol{\psi}}_2\right)\left(\sum_{k=1}^{i}\boldsymbol{y}_k - i\hat{\boldsymbol{\psi}}_2\right)'.$$

Next, we confirm unbiasedness of each estimator and derive their MSEs. First, we check the unbiasedness of the $\hat{\boldsymbol{\psi}}$ estimators $\hat{\boldsymbol{\psi}}_1$ and $\hat{\boldsymbol{\psi}}_2$. For $\hat{\boldsymbol{\psi}}_1$, we can easily confirm that $E[\hat{\boldsymbol{\psi}}_1] = \boldsymbol{\psi}$ from equation (3.7). Then,

$$E\left[\sum_{i=1}^{T-1}\sum_{k=1}^{i} i\boldsymbol{Y}_k\right] = \sum_{i=1}^{T-1} E\left[\sum_{k=1}^{i} i\boldsymbol{Y}_k\right] = \sum_{i=1}^{T-1} i^2\boldsymbol{\psi},$$

so that $E[\hat{\boldsymbol{\psi}}_2] = \boldsymbol{\psi}$. Given that unbiasedness is ensured, MSE coincides with the variance of the estimator. First, from (3.7) and (3.9),

$$V[\boldsymbol{Y}_1 + \cdots + \boldsymbol{Y}_{T-1}] = \sum_{k=1}^{T-1} V[\boldsymbol{Y}_k] + 2\sum_{k=1}^{T-2}\mathrm{Cov}(\boldsymbol{Y}_k, \boldsymbol{Y}_{k+1}) = 2\Sigma,$$

and therefore

$$V_1 := \mathrm{MSE}(\hat{\boldsymbol{\psi}}_1) = V[\hat{\boldsymbol{\psi}}_1] = \frac{2}{(T-1)^2}\Sigma. \tag{3.16}$$

Next, for $\hat{\boldsymbol{\psi}}_2$ because

$$\begin{aligned}
&V\left[\sum_{i=1}^{T-1}\sum_{k=1}^{i} i\boldsymbol{Y}_k\right] \\
=&\sum_{k=1}^{T-1} V\left[\frac{(k+T-1)(T-k)}{2}\boldsymbol{Y}_k\right] \\
&\quad + 2\sum_{k=1}^{T-2}\mathrm{Cov}\left(\frac{(k+T-1)(T-k)}{2}\boldsymbol{Y}_k, \frac{(k+T)(T-k-1)}{2}\boldsymbol{Y}_{k+1}\right) \\
=&2(T-1)^2\Sigma + \sum_{k=1}^{T-2}\frac{(k+T-1)^2(T-k)^2}{2}\Sigma \\
&\quad - \sum_{k=1}^{T-2}\frac{(k+T-1)(T-k)(k+T)(T-k-1)}{2}\Sigma \\
=&\left(2(T-1)^2 + \sum_{k=1}^{T-2} k(k+T-1)(T-k)\right)\Sigma \\
=&\frac{1}{12}T(T-1)(T+1)(3T-2)\Sigma
\end{aligned}$$

using (3.7) and (3.9), we obtain

$$V_2 := \mathrm{MSE}(\hat{\boldsymbol{\psi}}_2) = V\left[\hat{\boldsymbol{\psi}}_2\right] = \frac{3(T+1)(3T-2)}{T(T-1)(2T-1)^2}\Sigma. \tag{3.17}$$

By (3.16) and (3.17), both unbiased estimators $\hat{\boldsymbol{\psi}}_1$ and $\hat{\boldsymbol{\psi}}_2$ are consistent estimators as $T \to \infty$ by the Chebyshev inequality and $\lim V_1 = \lim V_2 = 0$. Furthermore, the estimators constructed by $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_{T-1}$ are normal distributions. Thus,

$$\hat{\boldsymbol{\psi}}_1 \sim \mathcal{N}_n(\boldsymbol{\psi}, V_1), \qquad \hat{\boldsymbol{\psi}}_2 \sim \mathcal{N}_n(\boldsymbol{\psi}, V_2).$$

Next, we confirm the unbiasedness of $\hat{\Sigma}_1$ and $\hat{\Sigma}_2$. First, for the unbiasedness of $\hat{\Sigma}_1$, using

$$\begin{aligned}
&\sum_{i=1}^{T-1}\left(\boldsymbol{Y}_i - \hat{\boldsymbol{\psi}}_1\right)\left(\boldsymbol{Y}_i - \hat{\boldsymbol{\psi}}_1\right)' \\
&= \sum_{i=1}^{T-1}\left(\boldsymbol{Y}_i - \boldsymbol{\psi}\right)\left(\boldsymbol{Y}_i - \boldsymbol{\psi}\right)' - (T-1)(\hat{\boldsymbol{\psi}}_1 - \boldsymbol{\psi})(\hat{\boldsymbol{\psi}}_1 - \boldsymbol{\psi})',
\end{aligned}$$

(3.7) and (3.16), we obtain

$$E\left[\hat{\Sigma}_1\right] = \frac{T(T-2)}{(T-1)^2}\Sigma.$$

For the unbiasedness of $\hat{\Sigma}_2$, we know that

$$\begin{aligned}
&E\left[\sum_{i=1}^{T-1}\left(\sum_{k=1}^{i}\boldsymbol{Y}_k - i\boldsymbol{\psi}\right)\left(\sum_{l=1}^{i}\boldsymbol{Y}_l - i\boldsymbol{\psi}\right)'\right] \\
&= \sum_{i=1}^{T-1}\left(\sum_{k=1}^{i}V[\boldsymbol{Y}_k] + 2\sum_{k=1}^{i-1}\mathrm{Cov}(\boldsymbol{Y}_k, \boldsymbol{Y}_{k+1})\right) \\
&= 2(T-1)\Sigma.
\end{aligned} \tag{3.18}$$

Then, it can be further transformed in to

$$\begin{aligned}
&\sum_{i=1}^{T-1}\left(\sum_{k=1}^{i}\boldsymbol{Y}_k - i\hat{\boldsymbol{\psi}}_2\right)\left(\sum_{k=1}^{i}\boldsymbol{Y}_k - i\hat{\boldsymbol{\psi}}_2\right)' \\
&= \sum_{i=1}^{T-1}\left\{\left(\sum_{k=1}^{i}\boldsymbol{Y}_k - i\boldsymbol{\psi}\right)\left(\sum_{k=1}^{i}\boldsymbol{Y}_k - i\boldsymbol{\psi}\right)' - i^2(\boldsymbol{\psi} - \hat{\boldsymbol{\psi}}_2)(\boldsymbol{\psi} - \hat{\boldsymbol{\psi}}_2)'\right\}.
\end{aligned}$$

Therefore, from (3.17) and (3.18), we obtain

$$E\left[\hat{\Sigma}_2\right] = \frac{1}{T-1}\left(2(T-1)\Sigma - \sum_{i=1}^{T-1} i^2 V[\hat{\boldsymbol{\psi}}_2]\right) = \frac{(5T-3)(T-2)}{2(T-1)(2T-1)}\Sigma.$$

Because biases exist in $\hat{\Sigma}_1$ and $\hat{\Sigma}_2$, we set

$$\hat{\Sigma}_1^* = \frac{(T-1)^2}{T(T-2)}\hat{\Sigma}_1, \quad \hat{\Sigma}_2^* = \frac{2(T-1)(2T-1)}{(5T-3)(T-2)}\hat{\Sigma}_2$$

for the bias-corrected estimators.

Hereafter, we estimate $\sigma_\varepsilon^2$ and $\sigma_\zeta^2$ using the estimates of $\boldsymbol{\psi}$ and $\Sigma$. For brevity, let $\hat{\psi}_i, \hat{\sigma}_{i,j}^2$ denote the components of the $\boldsymbol{\psi}$ and $\Sigma$ estimators obtained thus far, respectively. First, because the off-diagonal components of $\Sigma$ depend only on $\sigma_\zeta^2$, we consider minimizing

$$\sum_{i=1}^{n}\sum_{j=i+1}^{n}\left(\hat{\sigma}_{i,j}^2 - \sigma_\zeta^2 \frac{\hat{\psi}_i\hat{\psi}_j}{\left(\sum_{x=1}^n \hat{\psi}_x\right)^2}\right)^2.$$

By differentiating the above equation with respect to $\sigma_\zeta^2$ and solving

$$2\sum_{i=1}^{n}\sum_{j=i+1}^{n}\frac{\hat{\psi}_i\hat{\psi}_j}{\left(\sum_{x=1}^n \hat{\psi}_x\right)^2}\left(\hat{\sigma}_{i,j}^2 - \sigma_\zeta^2 \frac{\hat{\psi}_i\hat{\psi}_j}{\left(\sum_{x=1}^n \hat{\psi}_x\right)^2}\right) = 0,$$

we get

$$\hat{\sigma}_\zeta^2 = \frac{\left(\sum_{x=1}^{n}\hat{\psi}_x\right)^2 \sum_{i=1}^{n}\sum_{j=i+1}^{n}\hat{\psi}_i\hat{\psi}_j\hat{\sigma}_{i,j}^2}{\sum_{i=1}^{n}\sum_{j=i+1}^{n}\hat{\psi}_i^2\hat{\psi}_j^2}. \tag{3.19}$$

Next, for diagonal components $\sigma_\varepsilon^2$,

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\sigma}_{i,i}^2 - \hat{\sigma}_\zeta^2 \frac{\hat{\psi}_i^2}{\left(\sum_{x=1}^n \hat{\psi}_x\right)^2}\right) \tag{3.20}$$

is obtained by finding $\sigma_\varepsilon^2$ that minimizes

$$\sum_{i=1}^{n}\left(\hat{\sigma}_{i,i}^2 - \sigma_\varepsilon^2 - \hat{\sigma}_\zeta^2 \frac{\hat{\psi}_i^2}{\left(\sum_{x=1}^n \hat{\psi}_x\right)^2}\right)^2$$

in the same way.

From the discussion above, we obtain an estimate of $\boldsymbol{\psi}$. Thus, we obtain the estimates of $\theta$ and $\boldsymbol{b}$ as

$$\hat{\theta} = \sum_{x=1}^{n} \hat{\psi}_x, \quad \hat{\boldsymbol{b}} = \frac{\hat{\boldsymbol{\psi}}}{\sum_{x=1}^{n} \hat{\psi}_x}. \tag{3.21}$$

Furthermore, from (3.3), we have

$$\hat{k}_t = \widehat{E(\kappa_t)} = \sum_{x=1}^{n} \left( \log q_{x,t} - \frac{\log q_{x,1} + \cdots + \log q_{x,T}}{T} \right) \tag{3.22}$$

for $t$ $(1 \leq t \leq T)$.

### 3.4. Probability points of probability densities of future mortality

In this section, we deduce the probability densities of $M_{x,t}$ and $\kappa_t$ for the future. The probability density functions include the unknown parameters $\theta$, $\boldsymbol{\psi}$, $\sigma_\varepsilon^2$, and $\sigma_\zeta^2$. Thus, we estimate the parameters as $\hat{\theta}$, $\hat{\boldsymbol{\psi}}$, $\hat{\sigma}_\varepsilon^2$, and $\hat{\sigma}_\zeta^2$ derived in the previous section using the observed values.

First, we consider the probability points of the predictive densities $M_{x,T+1}$ and $\kappa_{T+1}$. Since we assume that the model holds from the past and the future, (3.5) and (3.6) hold for $t \geq T$ as well. Then, since

$$\begin{aligned} &M_{x,T+1} | M_{x,T} \sim \mathcal{N}\left( M_{x,T} + \psi_x, 2\left( \frac{\psi_x}{\sum_{x=1}^{n} \psi_x} \right)^2 \sigma_\zeta^2 + 2\sigma_\varepsilon^2 \right), \\ &\kappa_{T+1} | \kappa_T \sim \mathcal{N}(\kappa_T + \theta, 2\sigma_\zeta^2). \end{aligned}$$

The predictive distributions are estimated as follows:

$$\begin{aligned} &\hat{f}_{M_{x,T+1}|\boldsymbol{m}} \sim \mathcal{N}\left( m_{x,T} + \hat{\psi}_x, 2\left( \frac{\hat{\psi}_x}{\sum_{x=1}^{n} \hat{\psi}_x} \right)^2 \hat{\sigma}_\zeta^2 + 2\hat{\sigma}_\varepsilon^2 \right), \\ &\hat{f}_{\kappa_{T+1}|\boldsymbol{m}} \sim \mathcal{N}(\hat{k}_T + \hat{\theta}, 2\hat{\sigma}_\zeta^2) \end{aligned}$$

by using the parameter estimates obtained in the previous section, where $\boldsymbol{m} = (\boldsymbol{m}_1, \ldots, \boldsymbol{m}_T)$. By using the above distributions, the lower and upper $\alpha/2\%$ points of $M_{x,T+1}$ are

$$m_{x,T} + \hat{\psi}_x \pm z_{\left(\frac{\alpha}{2}\right)} \sqrt{2\left( \frac{\hat{\psi}_x}{\sum_{x=1}^{n} \hat{\psi}_x} \right)^2 \hat{\sigma}_\zeta^2 + 2\hat{\sigma}_\varepsilon^2}.$$

The interval covers $(1-\alpha)\%$ of the distribution. Similarly, for $\kappa_{T+1}$, we have

$$\hat{k}_T + \hat{\theta} \pm z_{\left(\frac{\alpha}{2}\right)} \sqrt{2\hat{\sigma}_\zeta^2},$$

where $z_{\left(\frac{\alpha}{2}\right)}$ denotes the upper $\alpha/2\%$ point of the standard normal distribution and $\hat{k}_T$ is estimated by (3.22) with $t = T$.

For forecast $h$ $(\geq 2)$ years ahead of time $T$, by summing both sides of (3.6) and the second equation in (3.5) from $T+1$ to $T+h$, respectively, we obtain

$$\begin{aligned} M_{x,T+h} &= M_{x,T} + h\psi_x + \left( \frac{\psi_x}{\sum_{x=1}^n \psi_x} (\zeta_{T+h} - \zeta_T) + (\varepsilon_{x,T+h} - \varepsilon_{x,T}) \right), \\ \kappa_{T+h} &= \kappa_T + h\theta + \zeta_{T+h} - \zeta_T, \end{aligned}$$

and thus, the conditional distributions are

$$\begin{aligned} M_{x,T+h}|M_{x,T} &\sim \mathcal{N}\left( M_{x,T} + h\psi_x, 2\left( \frac{\psi_x}{\sum_{x=1}^n \psi_x} \right)^2 \sigma_\zeta^2 + 2\sigma_\varepsilon^2 \right), \\ \kappa_{T+h}|\kappa_T &\sim \mathcal{N}(\kappa_T + h\theta, 2\sigma_\zeta^2). \end{aligned} \tag{3.23}$$

Similarly, by estimating

$$\hat{f}_{M_{x,T+h}|\boldsymbol{m}} \sim \mathcal{N}\left( m_{x,T} + h\hat{\psi}_x, 2\left( \frac{\hat{\psi}_x}{\sum_{x=1}^n \hat{\psi}_x} \right)^2 \hat{\sigma}_\zeta^2 + 2\hat{\sigma}_\varepsilon^2 \right), \tag{3.24}$$

$$\hat{f}_{\kappa_{T+h}|\boldsymbol{m}} \sim \mathcal{N}(\hat{k}_T + h\hat{\theta}, 2\hat{\sigma}_\zeta^2) \tag{3.25}$$

for the predictive distribution $h$ years ahead, the lower and upper $\alpha/2\%$ of $M_{x,T+h}$ are

$$m_{x,T} + h\hat{\psi}_x \pm z_{\left(\frac{\alpha}{2}\right)} \sqrt{2\left( \frac{\hat{\psi}_x}{\sum_{x=1}^n \hat{\psi}_x} \right)^2 \hat{\sigma}_\zeta^2 + 2\hat{\sigma}_\varepsilon^2}. \tag{3.26}$$

Similarly, for $\kappa_{T+h}$, we have

$$\hat{k}_T + h\hat{\theta} \pm z_{\left(\frac{\alpha}{2}\right)} \sqrt{2\hat{\sigma}_\zeta^2}. \tag{3.27}$$

## §4. A simulation study

### 4.1. Unbiasedness and MSE of the proposed estimator

In this section, we check the unbiasedness and theoretical MSE derived from the proposed estimators. We also evaluate the performance of the computational estimate for the quasi-likelihood function using an optimization function in R. The data were generated under the following settings and verified using Monte Carlo simulations:

- Number of Monte Carlo replications: $MC = 2000$,
- $T = 70$,
- $\boldsymbol{\psi} = (-0.02, -0.03)'$, $\sigma_{\boldsymbol{\varepsilon}}^2 = 0.001$, $\sigma_\zeta^2 = 0.1$.

The results of the simulation settings are shown below. First, the unbiasedness results for $\boldsymbol{\psi}$ are listed in Table 1. We observe that both estimators are unbiased, as are the theoretical values. $\hat{\boldsymbol{\psi}}_{\text{opt}}$ in Table 1 is an estimate of the optimization function in R.

Table 1: Simulation results of bias for $\boldsymbol{\psi}$.

| | Bias |
|---|---|
| $\hat{\boldsymbol{\psi}}_1$ | $(6.244 \cdot 10^{-5}, 8.036 \cdot 10^{-5})'$ |
| $\hat{\boldsymbol{\psi}}_2$ | $(6.053 \cdot 10^{-5}, 7.996 \cdot 10^{-5})'$ |
| $\hat{\boldsymbol{\psi}}_{\text{opt}}$ | $(8.435 \cdot 10^{-5}, 9.122 \cdot 10^{-5})'$ |

The MSE results are listed in Table 2. The estimated results of the MSEs are close to the theoretical MSEs as shown in Table 3, indicating that both methods are correctly estimated.

Table 2: Estimated MSE for $\boldsymbol{\psi}$.

| | MSE |
|---|---|
| $\hat{\boldsymbol{\psi}}_1$ | $\begin{pmatrix} 6.912 \cdot 10^{-6} & 9.787 \cdot 10^{-6} \\ 9.787 \cdot 10^{-6} & 1.523 \cdot 10^{-5} \end{pmatrix}$ |
| $\hat{\boldsymbol{\psi}}_2$ | $\begin{pmatrix} 8.035 \cdot 10^{-6} & 1.134 \cdot 10^{-5} \\ 1.134 \cdot 10^{-5} & 1.749 \cdot 10^{-5} \end{pmatrix}$ |
| $\hat{\boldsymbol{\psi}}_{\text{opt}}$ | $\begin{pmatrix} 9.765 \cdot 10^{-6} & 1.178 \cdot 10^{-5} \\ 1.178 \cdot 10^{-5} & 1.915 \cdot 10^{-5} \end{pmatrix}$ |

Table 3: Theoretical MSE for $\boldsymbol{\psi}$.

| | Theoretical MSE |
|---|---|
| $\hat{\boldsymbol{\psi}}_1$ | $\begin{pmatrix} 7.141 \cdot 10^{-6} & 1.008 \cdot 10^{-5} \\ 1.008 \cdot 10^{-5} & 1.554 \cdot 10^{-5} \end{pmatrix}$ |
| $\hat{\boldsymbol{\psi}}_2$ | $\begin{pmatrix} 8.070 \cdot 10^{-6} & 1.139 \cdot 10^{-5} \\ 1.139 \cdot 10^{-5} & 1.756 \cdot 10^{-5} \end{pmatrix}$ |

The biases and MSEs of $\sigma_\varepsilon^2$ and $\sigma_\zeta^2$ are presented in Tables 4 and 5. Here, Method 1 in the tables indicates the results calculated from $\hat{\boldsymbol{\psi}}_1$ and $\hat{\Sigma}_1^*$, and Method 2 is calculated from $\hat{\boldsymbol{\psi}}_2$ and $\hat{\Sigma}_2^*$. Method 3 is based on the optimization function of R. Additionally, the (scalar) MSE of $\hat{\boldsymbol{\psi}}$ in Table 5 is defined as $\text{Tr}[\text{MSE}(\hat{\boldsymbol{\psi}})]$. It can be seen that both results are estimated more correctly than when using the optimization function. The results also show that the model can be estimated correctly using a simple method. Because the former can be solved algebraically, the computational time is instant, whereas the optimization function generally requires a considerable computation time.

Table 4: Simulation results of estimated bias for $\boldsymbol{\psi}, \sigma_\varepsilon^2, \sigma_\zeta^2$.

| | Bias | | |
|---|---|---|---|
| | $\|\|\boldsymbol{\psi}\|\|$ | $\sigma_\varepsilon^2$ | $\sigma_\zeta^2$ |
| Method 1 | $1.017 \cdot 10^{-5}$ | $5.264 \cdot 10^{-5}$ | $1.008 \cdot 10^{-3}$ |
| Method 2 | $1.002 \cdot 10^{-5}$ | $3.435 \cdot 10^{-5}$ | $1.602 \cdot 10^{-4}$ |
| Method 3 | $1.242 \cdot 10^{-5}$ | $-1.095 \cdot 10^{-2}$ | $8.895 \cdot 10^{-2}$ |

Table 5: Simulation results of estimated MSE for $\boldsymbol{\psi}, \sigma_\varepsilon^2, \sigma_\zeta^2$.

| | MSE | | |
|---|---|---|---|
| | $\boldsymbol{\psi}$ | $\sigma_\varepsilon^2$ | $\sigma_\zeta^2$ |
| Method 1 | $2.214 \cdot 10^{-5}$ | $4.094 \cdot 10^{-7}$ | $4.372 \cdot 10^{-4}$ |
| Method 2 | $2.553 \cdot 10^{-5}$ | $8.836 \cdot 10^{-7}$ | $1.116 \cdot 10^{-3}$ |
| Method 3 | $2.892 \cdot 10^{-5}$ | $1.384 \cdot 10^{-4}$ | $7.930 \cdot 10^{-3}$ |

### 4.2. Comparison with SVD

We compare our parameter estimation with the SVD method used in the LC model. Since the LC model is 2 stage modeling, the data structure is different from a single random walk model by us and King and Girosi. In fact, $\sigma_\zeta^2$ is not used in the SVD estimation in the LC model, while our and their models include it in the data generating process. Therefore, it is difficult to compare them directly. However, we generate the data based on the SVD method by (2.1), and then compare the SVD method and our model. We set $n = 2$ and $T = 70$. For the comparison, let define true parameters as

$$\boldsymbol{a} = \begin{pmatrix} 0.4 \\ 0.6 \end{pmatrix}, \quad \boldsymbol{b} = \begin{pmatrix} 0.4 \\ 0.6 \end{pmatrix}, \quad k_{t+1} = \theta + k_t \quad (k = 1, \ldots, T-1),\ k_1 = 138,$$

where $\theta = -4$. Note that $\boldsymbol{b}$ and $\boldsymbol{k}$ satisfy (2.3). We generate the logarithm $\boldsymbol{m}_t$ of mortalities $\boldsymbol{q}_t$ $(t = 1, ..., T)$ using (2.1) with $\sigma_\varepsilon^2 = 0.001$.

The SVD method outputs the estimates of $\boldsymbol{a}$, $\boldsymbol{b}$, and $\boldsymbol{k}$ using the generated logarithm $\boldsymbol{m}_t$ of the mortalities $(1 \leq t \leq T)$. As mentioned in [10], the parameter $\theta$ in the forecasting part is estimated by

$$\hat{\theta} = \frac{\hat{k}_T - \hat{k}_1}{T - 1},$$

and then $\hat{\boldsymbol{\psi}}$ is obtained by (3.1) and $\hat{\boldsymbol{b}}$. In our parameter estimation, we obtain the estimate of $\boldsymbol{\psi}$ by (3.15) using the same generated date $\boldsymbol{m}_t$ $(1 \leq t \leq T)$, and then the estimates of $\boldsymbol{b}$ and $\theta$ by (3.21). The estimate of $\boldsymbol{a}$ is the same as SVD as in (2.7). Finally, $\boldsymbol{k}$ is estimated by (3.22).

Now, we obtain a sequence of estimates using a Monte Carlo simulation with $MC = 2000$ replications. Table 6 is the results for the estimated scalar MSE of these estimates.

Table 6: The estimated scalar MSEs for parameters.

| | $\boldsymbol{\psi}$ | $\boldsymbol{b}$ | $\boldsymbol{k}$ | $\theta$ |
|---|---|---|---|---|
| SVD | $4.430 \times 10^{-7}$ | $2.342 \times 10^{-9}$ | 0.132 | $7.865 \times 10^{-7}$ |
| our estimate | $9.522 \times 10^{-7}$ | $3.130 \times 10^{-8}$ | 0.138 | $9.486 \times 10^{-7}$ |

Even if the data is generated by the SVD model, our model has extremely close performance to SVD. Furthermore, our estimate is easy to calculate by (3.15), (3.21), and (3.22) without calculating eigen values using in SVD.

### 4.3. Forecasts

In this section, we verify the accuracy of the probability points obtained by (3.26) and (3.27) in the previous section by using Monte Carlo simulations under the following settings:

- Number of Monte Carlo replications: $MC = 1000$,
- $T = 50, 100, 500, \quad h = 10$,
- $\boldsymbol{\psi} = (-0.02, -0.03)', \quad \sigma^2_{\boldsymbol{\varepsilon}} = 0.001$, $\sigma^2_{\zeta} = 0.1$.

$T = 500$ is unrealistic because large-scale data are not available in practice. To evaluate the performance of our estimates, we created a virtual situation. The other settings are particularly useful in the current situation, since only approximately 70 years of mortality data are available. We must generate both values in the observed period and the future period of a random process. First, we generate the observed values in the random process from (3.5) with random errors in (3.4), and the true value of $\theta$ using (3.1) by $\{\kappa_t^{(j)}\}_{1\le t\le T}$ for the $j$th Monte Carlo replication. Note that we do not use assumption (3.10) when we generate the sequence because this assumption is only used to estimate the parameters in the random process using the quasi-likelihood function. The original random process does not assume that (3.10). In fact, we generate $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{T-1}$ with (3.7). The parameter estimates of $\theta$, $\psi_x$, $\sigma^2_\zeta$, $\sigma^2_\varepsilon$ in (3.26) and (3.27) are conducted using $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{T-1}$. We set the initial value $\boldsymbol{m}_1 = (1, 2)'$ that is required to calculate $m_{x,T}$ and $\hat{k}_T$ in (3.26) and (3.27) in our simulation, which is independent of the length of the predictive intervals. A future value $\kappa^{(j)}_{T+h}$ has a conditional distribution given $\kappa^{(j)}_1, \ldots, \kappa^{(j)}_T$. To obtain the distribution of $\kappa^{(j)}_{T+h}$, we generated $N = 10000$ samples from the conditional distribution. In fact, we generate the samples using (3.5) and not directly from (3.23). Let $\kappa_{T+h}^{(l,j)}$ be the $l$th ($1 \le l \le N$) of the samples. To obtain the upper 1% point, let $U_{0.01}\left(\kappa_{T+h}^{(j)}\right)$ denote the 9900th order value when we set $\kappa_{T+h}^{(1,j)}, \ldots, \kappa_{T+h}^{(N,j)}$ in ascending order for each $j$. Using our notation, we provide the upper and lower $\alpha/2\%$ in (3.27). If we set $\alpha = 2$, we obtain the upper 1%. We generated $N = 10000$ samples, and $U_{0.01}\left(\kappa_{T+h}^{(j)}\right)$ may be regarded as the true 99% percent point of $\kappa_{T+h}^{(j)}$. For each $j$, we estimate $U_{0.01}\left(\kappa_{T+h}^{(j)}\right)$ by using our estimate. Let $\widehat{U_{0.01}\left(\kappa_{T+h}^{(j)}\right)}$ be an estimate of the second formula in (3.27). This is a point estimate of $U_{0.01}\left(\kappa_{T+h}^{(j)}\right)$ given $\kappa_1^{(j)}, \ldots, \kappa_T^{(j)}$.

$U_{0.01}\left(\kappa_{T+h}^{(j)}\right)$ and $\widehat{U_{0.01}\left(\kappa_{T+h}^{(j)}\right)}$ are defined for each $j$ ($1 \le j \le MC$). The values depend on $\kappa_1^{(j)}, \ldots, \kappa_T^{(j)}$ because the parameters in $\widehat{U_{0.01}\left(\kappa_{T+h}^{(j)}\right)}$ are estimated using the observed data generated for each $j$. Furthermore, the estimate has variation in the estimation since $\widehat{U_{0.01}\left(\kappa_{T+h}^{(j)}\right)}$ is a point estimate. To validate the accuracy of the estimate, we evaluated it using a Monte Carlo setting. We evaluated the upper 1% of the points

$$U_{0.01}\left(\kappa_{T+h}\right) \simeq \frac{1}{MC}\sum_{j=1}^{MC} U_{0.01}\left(\kappa_{T+h}^{(j)}\right)$$

by

$$\widehat{U_{0.01}\left(\kappa_{T+h}\right)} = \frac{1}{MC}\sum_{j=1}^{MC} \widehat{U_{0.01}\left(\kappa_{T+h}^{(j)}\right)}.$$

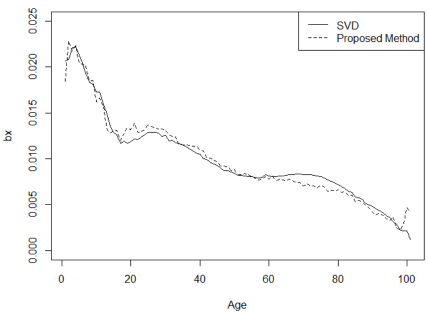
The same procedure was applied for $\boldsymbol{M}_t$. In Table 7 only the upper points are listed. A comparison of the results shows that the estimates are accurate for each $T$ setting.

Table 7: Upper points of $\boldsymbol{M}_t$ and $\kappa_t$.

| | $T = 50$ | $T = 100$ | $T = 500$ |
|---|---|---|---|
| $U_{0.01}\left(\kappa_{T+h}\right)$ | 0.1618 | $-0.0365$ | $-1.6308$ |
| $\widehat{U_{0.01}\left(\kappa_{T+h}\right)}$ | 0.1610 | $-0.0373$ | $-1.6308$ |
| $U_{0.01}\left(\boldsymbol{M}_{T+h}\right)$ | $(0.03505, 0.1348)'$ | $(-0.06552, 0.03408)'$ | $(-0.8638, -0.7643)'$ |
| $\widehat{U_{0.01}\left(\boldsymbol{M}_{T+h}\right)}$ | $(0.03456, 0.1344)'$ | $(-0.06596, 0.03374)'$ | $(-0.8638, -0.7644)'$ |

## §5. Real data analysis

Figures 1 and 2 show the estimation results of $b_x$ and $k_t$ obtained by SVD, and $\hat{\boldsymbol{\psi}}_1$ in (3.14). We propose using mortality data [11] from 1947 to 2020. For our estimates, we estimate the values of $b_x$ using (3.21) and the values of $k_t$ using (3.22).



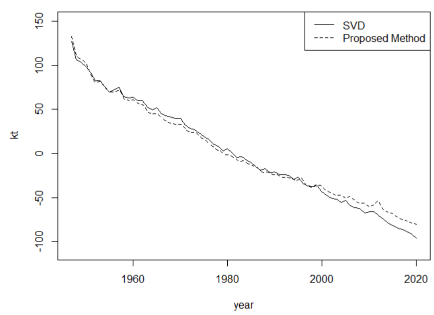




Figure 1: Estimation results for $b_x$. Figure 2: Estimation results for $k_t$.

Because both have similar values, the proposed estimation method under our assumption is compatible with the model proposed by Lee and Carter. Our estimate does not require complex computation, whereas the method proposed by Lee and Carter uses SVD. As we can see, the estimate $\hat{\boldsymbol{\psi}}_1$ of $\boldsymbol{\psi}$ is the simple average of the observed values. Even when we use $\hat{\boldsymbol{\psi}}_2$, this value is a weighted average. For $k_t$, we did not use the estimate of $\boldsymbol{\psi}$. We calculate from the mortality data, as in (3.22). Thus, we conclude that we have given an appropriate and simple method to estimate the parameters that were not discussed by Girosi and King, and have dissolved the problem of inconsistency by Lee and Carter.

Next, we show the upper and lower 1% probability points of $\boldsymbol{M}_t$ and $\kappa_t$ obtained by applying the real data to (3.26) and (3.27) for the age groups of 20 to 29 and from 70 to 79, respectively. The data are obtained from the same source, but we conduct our estimation for two different datasets: the young group and the older group independently. In our notation, age $x$ ranges from 1 to $n$, where $n$ is the final age in the life table. The model can be easily applied to 20 to 29 age groups by corresponding to 1 to $n = 10$ as well as 70 to 79.

Although $\boldsymbol{M}_t$ is $n = 10$ dimension data, the results of the prediction by (5.1) for ages 20 and 70 are shown on behalf of their groups in Figures 3 and 4, respectively. For $\boldsymbol{M}_t$, we have the observed values $\boldsymbol{m}_t$ ($1 \leq t \leq T$) shown as the solid line on the left side of the vertical line in Figures 3 and 4. For the prediction of $\boldsymbol{M}_t$, we use the expectation of $\boldsymbol{M}_t$. From (3.24), we have

$$\hat{E}(M_{x,T+h}|M_{x,t} = m_{x,t}\ (1 \le t \le T)) = m_{x,T} + h\hat{\psi}_x. \tag{5.1}$$

The solid and dotted lines on the right side of the vertical line in Figures 3 and 4 represent the prediction of $\boldsymbol{M}_t$ for $t$ $(T+1 \le t \le T+h)$ and their upper and lower probability points by (3.26), respectively.

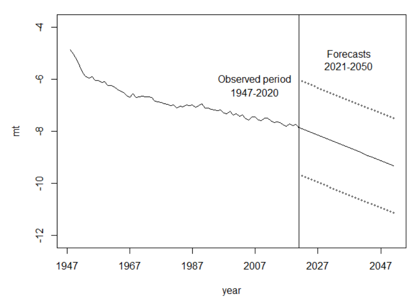



Figure 3: The observed values for $\boldsymbol{m}_t$ $(1 \le t \le T)$ and the estimated mean of $\boldsymbol{M}_t$ with the upper and lower 1% points of $\boldsymbol{M}_t$ at age 20 for $t$ $(T+1 \le t \le T+h)$.

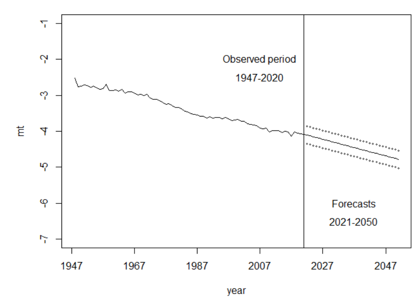



Figure 4: The observed values for $\boldsymbol{m}_t$ $(1 \le t \le T)$ and the estimated mean of $\boldsymbol{M}_t$ with the upper and lower 1% points of $\boldsymbol{M}_t$ at age 70 for $t$ $(T+1 \le t \le T+h)$.

For $\kappa_t$, we estimate $k_t = E(\kappa_t)$ for the observed and future periods. For the observed period, we estimated the values of $k_t$ using (3.22). The solid line on the left part of the vertical line in Figures 5 and 6 represents the estimated result for $k_t$ for $t$ $(1 \le t \le T)$. The prediction of $\kappa_t$ is given by

$$\hat{k}_{T+h} = \widehat{E(\kappa_{T+h})} = \hat{k}_T + h\hat{\theta}$$

in equation (3.25).

Because Lee and Carter applied a stochastic model in $\kappa_t$ only for the prediction part following ARIMA(0,1,0) and estimated the parameters in $\boldsymbol{M}_t$ using SVD which is not a stochastic model, they could only draw probability points for $\kappa_t$ and did not give probability points for $\boldsymbol{M}_t$. Our model has a standard error for $\boldsymbol{M}_t$ because it is constructed using the integrated stochastic model, which is different from the original LC model. In other words, we propose a method to obtain the variances of $\sigma_\varepsilon^2$ and $\sigma_\zeta^2$ that allows the probability points of $\kappa_t$ and $\boldsymbol{M}_t$. Note, however, that, as the results show, the width of each interval of the probability points is considerably different. This phenomenon is discussed in the final section.

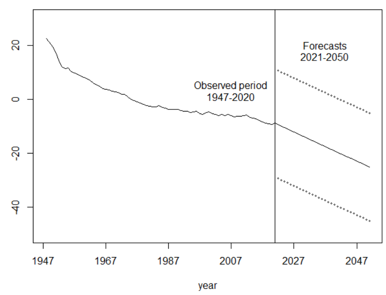



Figure 5: Estimated and predicted results for $k_t$ ($1 \leq t \leq T+h$) and for the upper and lower 1% points of $\kappa_t$ in the 20s ($T+1 \leq t \leq T+h$).

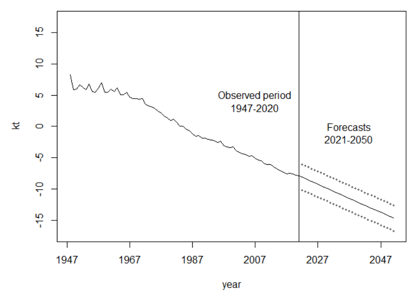



Figure 6: Estimated and predicted results for $k_t$ ($1 \leq t \leq T+h$) and for the upper and lower 1% points of $\kappa_t$ in the 70s ($T+1 \leq t \leq T+h$).

## §6. Discussion and future study

In this study, we propose a two-step estimation method to obtain $\sigma_\varepsilon^2$ and $\sigma_\zeta^2$ from the estimated values of $\boldsymbol{\psi}$ and $\Sigma$ using the quasi-likelihood function. An alternative approach is to estimate $\sigma_\varepsilon^2, \sigma_\zeta^2$, and $\boldsymbol{\psi}$ simultaneously using the quasi-likelihood function in (3.13). This is because we assumed $\boldsymbol{\varepsilon}_0 = \mathbf{0}$ and $\zeta_0 = 0$ to derive the likelihood function. A further extension is to estimate the unknown parameters from the exact likelihood function derived from Equation (3.6) as a multivariate time-series model. This would be a type of VAR or VMA model, but the estimation is generally difficult to calculate from the likelihood of a multivariate time-series model. Intensive consideration of parameter estimation is required if we treat the exact likelihood function. In fact, estimations for VARMA have identification difficulties (see e.g. [12]) and require much computational difficulty (see e.g. [13]). To conquer the problems approaches relying on Bayesian methods are used (see e.g. [14] and [15]). For example, Chang and Shi [14] applied a Bayesian factor-augmented approach. In general, a Bayesian approach requires many parameters and intensive computation. As we discussed, our approach is a very simple estimation method. The estimation ends with simple algebraic computation because it is based on the simple original LC model.

Additionally, in the future prediction of $\boldsymbol{M}_t$, the value of $M_{x,T+h}$ varies depending on the estimated values of $\psi_x$ as in (5.1). Note, however, that the estimates applied to real data performed in Section 5 show a decreasing trend for most ages. $\boldsymbol{\psi}$ is generally estimated to have negative values. The estimated values are negative because $\boldsymbol{\psi}$ is defined as the mean of the difference

$\boldsymbol{M}_t - \boldsymbol{M}_{t-1}$ by (3.7). Thus, if the estimated values of $\psi_x$ and $\psi_{x'}$ in $\boldsymbol{\psi}$ of age $x$ and $x'$ with $x < x'$ have a relation $\hat{\psi}_x > \hat{\psi}_{x'}$, then a phenomenon in which the mortality rate of age $x$ exceeds that of age $x'$ could occur in the future. As pointed out by Girosi and King, this is a result of the assumption that $\kappa_t$ behaves linearly. As can be seen from equations (2.6) and (3.5), $\boldsymbol{M}_t$ is only influenced by $\kappa_t$ with respect to time $t$; therefore, it is necessary to examine the validity of the LC model itself. As an approach to conquering the problem of linearity, Miyata and Matsuyama [16] mentioned the development of neural network approaches such as the NN-based generalization of the LC model using a fully connected network (see [17]). However, Miyata and Matsuyama [16] mentioned in their paper that neural network approaches are hard to interpret results if the methods are a single-stage estimation. Miyata and Matsuyama [16] proposed a method to solve their concern, however, the estimation results in their estimation (see Figure 5 in their paper) do not have similarity to the estimation result by the original LC model. We do not mention that their result is strange, but the results are far from the original LC model. We might interpret that Miyata and Matsuyama [16] could capture the phenomenon that the LC model does not.

We also point out another issue found through the real data analysis performed in Section 5. In the real data analysis, we divided the dataset into young and elderly people. We applied the proposed method to two independent datasets. Table 8 shows the results of the estimates of $\sigma_\varepsilon^2$ and $\sigma_\zeta^2$ for the young aged 20 to 29 and the elderly aged 70 to 79, respectively.

Table 8: Estimated results of $\sigma_\varepsilon^2$ and $\sigma_\zeta^2$.

| | $\hat{\sigma}_\varepsilon^2$ | $\hat{\sigma}_\zeta^2$ |
|---|---|---|
| young people | 0.0031 | 30.1310 |
| elderly people | 0.0012 | 0.3142 |

As it can be seen from the results, the estimation result of $\sigma_\zeta^2$ for the younger age group is much larger than the one for the elder age group. This is because of the original assumption that we have a common variance $\sigma_\zeta^2$ across ages, and it is an undesirable result that the estimation results are different for each age. As we know, many extensions of the LC model were proposed. For example, Fung et al. [18] weakened the assumption of shared variance. That is, they assume a heteroscedasticity model. Our main target is estimation methods for the original LC model. An extension of our approach to other models such as the models with heteroscedasticity is required in future.

## Acknowledgments

The authors thank the anonymous reviewers and the editors of the journal for their helpful and insightful suggestions.

## References

[1] R. D. Lee and L. R. Carter, "Modeling and forecasting U.S. mortality," *Journal of the American Statistical Association*, vol. **87**, no. 419, pp. 659–672, 1992.

[2] R. D. Lee, "The Lee-Carter method for forecasting mortality, with various extensions and applications," *North American Actuarial Journal*, vol. **4**, no. 1, pp. 80–91, 2000.

[3] H. Booth, J. Maindonald, and L. Smith, "Applying Lee-Carter under conditions of variable mortality decline," *Population Studies*, vol. **56**, no. 3, pp. 325–336, 2002.

[4] A. E. Renshaw and S. Haberman, "Lee-Carter mortality forecasting: A parallel generalized linear modelling approach for England and Wales mortality projections," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. **52**, no. 1, pp. 119–137, 2003.

[5] A. E. Renshaw and S. Haberman, "A cohort-based extension to the Lee-Carter model for mortality reduction factors," *Insurance: Mathematics and Economics*, vol. **38**, no. 3, pp. 556–570, 2006.

[6] S. Tuljapurkar, N. Li, and C. Boe, "A universal pattern of mortality change in the G7 countries," *Nature*, vol. **405**, pp. 789–792, 2000.

[7] J. R. Wilmoth, "Computational methods for fitting and extrapolating the Lee-Carter model of mortality change," Tech. Rep. 3, Berkeley, 1993.

[8] L. Siu-Hang, H. Mary, and T. Ken, "Uncertainty in mortality forecasting: An extension to the classical Lee-Carter approach," *ASTIN Bulletin*, vol. **39**, pp. 137–164, 2009.

[9] D. Harjani, S. Nurrohmah, and M. Novita, "Performance evaluation of the Bühlmann credibility approach in predicting mortality rates," *Journal of Physics: Conference Series*, vol. **1725**, Jan. 2021. 012095.

[10] F. Girosi and G. King, "Understanding the Lee-Carter mortality forecasting method." `https://gking.harvard.edu/files/abs/lc-abs.shtml`, 2007.

[11] National Institute of Population and Social Security Research, "Japanese mortality database." `http://www.ipss.go.jp/p-toukei/JMD/index.asp`, 2022.

[12] C. Gouriéroux, C. Monfort, and J. P. Renne, "Identification and estimation in non-fundamental structural VARMA models," *Review of Economic Studies*, vol. **87**, no. 4, pp. 1915–1953, 2020.

[13] J. C. Chan and E. Eisenstat, "Efficient estimation of Bayesian VARMAs with time-varying coefficients," *Journal of Applied Econometrics*, vol. **32**, no. 7, pp. 1277–1297, 2017.

[14] L. Chang and Y. Shi, "Mortality forecasting with a spatially penalized smoothed VAR model," *ASTIN Bulletin: The Journal of the IAA*, vol. **51**, no. 1, pp. 161–189, 2021.

[15] Y. Lu and D. Zhu, "Modelling mortality: A Bayesian factor-augmented VAR (FAVAR) approach," *ASTIN Bulletin: The Journal of the IAA*, vol. **53**, no. 1, pp. 29–61, 2023.

[16] A. Miyata and N. Matsuyama, "Extending the Lee-Carter model with variational autoencoder: A fusion of neural network and Bayesian approach," *ASTIN Bulletin: The Journal of the IAA*, vol. **52**, no. 3, pp. 789–812, 2022.

[17] R. Richman and M. Wüthrich, "A neural network extension of the Lee-Carter model to multiple populations," *Annals of Actuarial Science*, vol. **15**, no. 2, pp. 346–366, 2021.

[18] M. Fung, G. Peters, and P. Shevchenko, "A unified approach to mortality modelling using state-space framework: characterisation, identification, estimation and forecasting," *Annals of Actuarial Science*, vol. **11**, no. 2, pp. 343–389, 2017.

Reo Kanazawa
Department of Applied Mathematics, Tokyo University of Science,
Kagurazaka 1-3, Shinjuku, Tokyo 162-8601, Japan
*E-mail*: `1421508@alumni.tus.ac.jp`

Takeshi Kurosawa
Department of Applied Mathematics, Tokyo University of Science,
Kagurazaka 1-3, Shinjuku, Tokyo 162-8601, Japan
*E-mail*: `tkuro@rs.tus.ac.jp`